STAT 380 – Spring 2019

Problems: Assignment 1

1. Consider events E, F and G and prove $E(F \cup G) = EF \cup EG$. [To prove two sets, say A and B, are equal you prove that $\omega \in A$ implies $\omega \in B$ and the other way around.]

   To show $A = B$ where A and B are sets you must prove that $x \in A$ implies $x \in B$ and that $x \in B$ implies $x \in A$. That is, you prove each set is a subset of the other. So suppose $x \in E(F \cup G)$. This is equivalent to the statement '$x \in E$ and ($x \in F$ or $x \in G$)'. It is a basic rule of logic that P and (Q or R) is logically equivalent to (P and Q) OR (P and R). So

   $$x \in E \text{ and } (x \in F \text{ or } x \in G) \iff (x \in E \text{ and } x \in F) \text{ OR } (x \in E \text{ and } x \in G)$$

   But

   $$x \in E \text{ and } x \in F \iff x \in EF$$

   and

   $$x \in E \text{ and } x \in G \iff x \in EG$$

   so

   $$x \in E(F \cup G) \iff (x \in EF) \text{ OR } (x \in EG) \iff x \in EF \cup EG.$$

2. In the game of craps what is the probability that the player wins? [A player rolls a pair of dice and wins if s/he rolls a 7 or an 11 on the first roll and loses if s/he rolls a 2, 3, or 12 on the first roll. If the player rolls any other number $j \in \{4, 5, 6, 8, 9, 10\}$ on the first roll then s/he continues rolling until s/he rolls a j (and wins) or a 7 (and loses). It is possible to do this problem without using infinite sums and to get full marks you need to do it that way.]

Let W be the event you win. Let F_j be the event you roll a j on the first toss. Let B be the event that the first roll after the first toss which is either a 7 or the same as the first toss is the same as the first toss.

The short cut way to do this problem is to recognize that

$$P(B|F_j) = \frac{P(\text{roll a } j)}{P(\text{ roll a 7 or a } j)}$$

for j one of 4, 5, 6, 8, 9, 10. Then

$$W = F_7 \cup F_{11} \cup \bigcup_{j \in \{4,5,6,8,9,10\}} BF_j$$

so

$$P(W) = P(F_7) + P(F_{11}) + \sum_{j \in \{4,5,6,8,9,10\}} P(B|F_j)P(F_j)$$

This works out to

$$\frac{6}{36} + \frac{2}{36} + \frac{3}{36}\frac{3/36}{6/36 + 3/36} + \cdots$$

which works out to

$$\frac{8}{36} + \frac{2}{36}\left(\frac{3^2}{9} + \frac{4^2}{10} + \frac{5^2}{11}\right) = \frac{244}{495} = .49292929\ldots$$

However: many students will find that way to be too subtle. Here is a different way. Let B_j be the event that the first toss is a j and the player wins. Then

$$W = F_7 \cup F_{11} \cup B_4 \cup B_5 \cup B_6 \cup B_8 \cup B_9 \cup B_{10}.$$

These events are disjoint; they are the different ways to win. So we compute the probabilities of each. The easy ones are

$$P(F_7) = \frac{6}{36} = \frac{1}{6}$$

and

$$P(F_{11}) = \frac{2}{36} = \frac{1}{18}.$$

We need to compute $P(B_j)$ for j any of the numbers 4, 5, 6, 8, 9, or 10. Fix j and let C_n be the event that the first toss is a j, tosses number 2 through $n-1$ are neither 7 nor j and toss n is j. The probability of C_n is

$$P(C_n) = P(F_j)(1 - P(F_j) - 1/6)^{n-2} P(F_j)$$

The first term comes from the fact that you have to start with a j. The last is the fact that on toss n you get j. There are then $n-2$ trials on which you don't get j and you don't get a 7; the latter has chance 1/6. Let $\theta = 1 - P(F_j) - 1/6$. Then

$$P(C_n) = \theta^{n-2} \left\{P(F_j)\right\}^2 .$$

Since

$$B_j = \cup_{n=2}^{\infty} C_n$$

and the C_n are pairwise disjoint we find

$$\begin{aligned} P(B_j) &= P(\cup_{n=2}^{\infty} C_n) \\ &= \left\{P(F_j)\right\}^2 (1 + \theta + \theta^2 + \cdots) \\ &= \frac{\left\{P(F_j)\right\}^2}{1-\theta} = \frac{\left\{P(F_j)\right\}^2}{P(F_j) + 1/6}. \end{aligned}$$

Adding up all these as before gives the previous answer.

3. Tossing a fair coin. What is the probability the first 4 tosses are HHHH? THHH? Then what is the probability you see the pattern THHH before you see the pattern HHHH?

The first two probabilities are 1/16. Students are tempted to think the third probability is 1/2. But think about a long string of tosses. Start at the left and continue until you first see the sequence HHHH. There are two possibilities: the first of those 4 Hs is the first toss you made, or it is not and the toss before that H is a T. In the latter case the sequence THHH happens before the sequence HHHH. So unless the first 4 tosses are H, the sequence THHH happens first. The chance the it happens first is 1 minus the chance the first 4 tosses are Heads. So the chance THHH comes first is 15/16.

4. Female Jo and Male Joe have brown eyes. Their mothers had blue eyes. Blue eyes are recessive. They have a daughter Flo with brown eyes who is having a child with an unnamed man with blue eyes. What is the chance the child has blue eyes? [From each parent a child inherits one of two *alleles* — in this case either a Brown allelle or a Blue allele. A person with two blue alleles has blue eyes; that is wht 'recessive' means; A person with one brown and one blue allele or two brown alleles has brown eyes.]

   Jo has a blue-eyed gene and a brown-eyed gene and so does Joe. Flo inherits one eye-colour gene from each parent so the 4 possibilities are equally likely. We are given that Flo has brown eyes so we have been told that one of the four possibilities has been ruled out. Of the remaining 3 possibilities, one is that both of Flo's genes are brown and the other 2 correspond to Flo having one blue, one brown gene. The conditional probability, given the observed colours, of two brown eye alleles, is 1/3; the other possibility, of one blue, one brown has conditional probability 2/3. If Flo has two brown eye genes then her child cannot have blue eyes. If she has one gene of each colour then the chance that she passes on the blue eye gene is 1/2. In the latter case the child has blue eyes. So the desired chance is

$$\frac{2}{3}\frac{1}{2} = \frac{1}{3}.$$

5. You are given a biased coin and you don't know θ, the probability the coin lands Heads. You toss it twice. If you get different results you take the result of the second toss. If you don't get different results you repeat the process. You are asked to show that the probability your final result is H is 1/2. Then think about the simpler sounding scheme where you toss the coin one toss at a time until you get two consecutive tosses which are different and then use the second of those two. So if you got HHT you would take T as the result. Does the simpler scheme work to make H and T equally likely?

As you might imagine the second scheme doesn't work. Suppose the probability of heads, θ, is very small. Then you will probably toss the coin a lot of times before you see Heads for the first time. In this case the second of the two tosses will be H. So the second scheme produces H unless you get H on the first toss which happens with chance only θ. So now think about the scheme suggested in part a. Let E_n be the event that the first $n-1$ times you tossed the coin twice the results were the same and then the nth time you tossed TH. The chance of this is

$$\left[\theta^2 + (1-\theta)^2\right]^{n-1} (1-\theta)\theta.$$

Let F_n be the event that the first $n-1$ times you tossed the coin twice the results were the same and then the nth time you tossed HT. Then

$$P(F_n) = \left[\theta^2 + (1-\theta)^2\right]^{n-1} \theta(1-\theta) = P(E_n).$$

The probability the scheme produces Heads is

$$P(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} P(F_n).$$

If you are willing to believe that the chance that you eventually get a trial where the two tosses are different then

$$1 = P(\cup_{n=1}^{\infty}(E_n \cup F_n) = \sum_{n=1}^{\infty}[P(E_n) + P(F_n)] = 2\sum_{n=1}^{\infty} P(E_n).$$

So the desired probability is $1/2$. Alternatively if you set

$$\phi = \theta^2 + (1-\theta)^2 = 2\theta^2 - 2\theta + 1$$

then

$$P(\cup_{n=1}^{\infty} E_n) = \theta(1-\theta)\sum_{n=1}^{\infty} \phi^{n-1} = \frac{\theta(1-\theta)}{1-\phi}$$

The latter is

$$\frac{\theta(1-\theta)}{2\theta - 2\theta^2} = \frac{1}{2}.$$

6. Suppose X is a non-negative integer valued rv and show that

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \ge n).$$

The second equality is easy because the terms match one by one. If

$$I_n = 1(n \le X)$$

then

$$\sum_{n=1}^{\infty} I_n = X$$

because: for each $n \le X$ the variable $I_n = 1$ while $n > X$ implies that $I_n = 0$. There are X integers n which are less than or equal to X and greater than or equal to 1; they are the integers $1, \ldots, X$. So

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(\sum_{n=1}^{\infty} I_n) \\
&= \sum_{n=1}^{\infty} \mathrm{E}(I_n) \\
&= \sum_{n=1}^{\infty} P(X \ge n)
\end{aligned}$$

The last line is because random variables whose only possible values are 0 and 1 have expected value equal to the probability that they are 1.

There are other approaches and most students will do the following, I guess, because this way is probably findable on the web:

$$\begin{array}{lcccccccc}
P(X>0) & = & P(X=1) & + & P(X=2) & + & P(X=3) & + & \cdots \\
P(X>1) & = & & & P(X=2) & + & P(X=3) & + & \cdots \\
P(X>2) & = & & & & & P(X=3) & + & \cdots \\
\vdots & \vdots & \cdots & & & & & &
\end{array}$$

Now add up each column: you get

$$1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots$$

so

$$\sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} nP(X = n) = \mathrm{E}(X).$$

7. Consider a population of 200 million people of whom 200 thousand have a certain condition. A test is available with the following properties. Assuming that a person has the condition the probability that the test detects the condition is 0.9. Assuming that a person does not have the condition the test detects (incorrectly) the condition with probability 0.001. A person is picked at random from the 200 million people and the test is administered.

   (a) What is the chance that the test detects the condition for this randomly selected person?

   Let C be the event that the selected person has the condition. Let D be the event that the test 'detects' the condition – correctly or otherwise. The information given is that

   $$P(C) = \frac{200,000}{200,000,000} = 0.001,$$
   $$P(D|C) = 0.9,$$

   and

   $$P(D|C^c) = 0.001.$$

   We are to compute $P(D)$ so we add

   $$P(D) = P(DC) + P(DC^c)$$

   Then we use $P(AB) = P(A|B)P(B)$ to get

   $$\begin{aligned} P(D) &= P(D|C)P(C) + P(D|C^c)P(C^c) \\ &= 0.9 \times 0.001 + 0.001 \times (1 - 0.001) \\ &= 0.001899, \end{aligned}$$

(b) Assuming that the condition is detected by the test for this randomly selected person what is the chance that the person has the condition?

Now we want $P(C|D)$. So we use Baye's theorem

$$
\begin{aligned}
P(C|D) &= \frac{P(CD)}{P(D)} \\
&= \frac{P(D|C)P(C)}{P(D)} \\
&= \frac{0.9 \times 0.001}{0.001899} \\
&= 0.4739336 \approx 0.474.
\end{aligned}
$$

Notice that this is less than 50%!

(c) A mandatory testing program is contemplated. If all 200 million are tested about how many positive results should be expected? Of these about how many will not have the condition?

The expected number of positive test results is

$$200,000,000 \times P(D) = 379,800.$$

The expected number of people with a positive test result who don't have the condition is

$$200,000,000 \times P(DC^c) = 199,800.$$

This is a result of considerable practical importance. Almost all medical tests, drug tests, and so on, have non-negligible false positive rates and false negative rates. In the example I asked you to do the sensitivity is 0.9; the false negative rate is 0.1. The specificity is 0.999 which means the false positive rate is 0.001. The *prevalence* is also 0.001 (200,000 out of 200,000,000) which is why that very small false positive rate is important. The same issues arise in *forensics*.

8. In this question I want you to use `R` to do a simulation. `R` is freely available at /https://www.r-project.org/. I want you to study the distribution of the t statistic when the population distribution is not normal.

Use the function `rgamma` built into `R` to generate samples $X_1, \ldots, X_n$ from the Gamma distribution with various values of the shape parameter. For each sample you generate compute the t-pivot

$$\frac{\sqrt{n}\left(\bar{X}-\mu\right)}{s}$$

where $\bar{X}$ and s are the usual sample means and standard deviation. Then for each of 4 different values of the shape parameter (I suggest 0.1, 1, 10, and 100) you should generate a large number of samples – say $M = 10,000$ – and compute the t-statistic for each sample. For each shape parameter prepare a histogram using `hist` with the argument `probability` set to true. Superimpose a plot of the t density with the appropriate number of degrees of freedom on each histogram. Try $n = 30$ and $n = 100$ and comment in your answer on what you learned about the rule of thumb that the central limit theorem is useful for $n \geq 30$.

I want the `R`-code you use to generate the data, compute the statistics and draw the histograms and curves mailed to me. Send me by e-mail a short paragraph containing your comments about the central limit theorem in this context.

The following `R`-code might be useful as an example. It creates a function to compute the t statistic for a vector `x` of data when the true (or hypothesized) mean is `mn`. Then it generates a matrix of random numbers from the Weibull distribution with shape 1/2. The matrix is 100 by 10,000. The line `tv = apply(utem,2,tst,mn=2)` computes the t-statistic for each column of that matrix using the true mean of 2. Then I draw the histogram and superimpose a standard normal curve (but you should use t curves).

```
tst = function(x,mn){ sqrt(length(x))*(mean(x)-mn)/sqrt(var(x))}
utem = matrix(rweibull(1000000,shape=0.5),nrow=100)
tv = apply(utem,2,tst,mn=2)
hist(tv,breaks=80,prob=T)
curve(dnorm(x),-4,4,add=T,col='red')
```

> I will not be producing a solution set for this question. However I notice that students often arrived at conclusions differing from mine. The code in the question makes 10,000

samples of size 100 and computes 10,000 t-statistics. The mean of the weibull distribution in question is 2. The code puts a normal curve on top. In your code you need 8 pictures and 8 matrices of data. For each shape you need a matrix with 100 rows and a matrix with only 20 rows.

Here are some things I sometimes see in the answers and some comments about the results when I did this problem:

- A number of students used only sample size 100.
- When the shape is 0.1 and the sample size is 30 the histogram of t statistics is very skewed to the left.
- Many students thought the t approximation was good in every case. This is certainly not true for $n = 30$ and shape $= 0.1$.
- Gamma distributions with small shape parameters are strongly skewed to the *right*. The corresponding t histograms are skewed to the *left*.
- If the population distribution is very skewed (like shape 0.1) then the rule of thumb is poor. Notice, though, that these histograms are not just about the normal approximation but *also* about the t distribution. You could compute histograms of sample means with normal curves on top to just look at CLT.

My code which you could run if you want.

```
tst = function(x,mn){
  sqrt(length(x))*(mean(x)-mn)/sqrt(var(x))
}

#
#  Generate a 100 by 10,000 matrix with shape =0.1.
#   This Gamma distribution has mean 0.1.
#
x100 = matrix(rgamma(1000000,shape=0.1),nrow=100)
#
#  Use the first 30 rows of x111 for the n=30
#
```

```
x30 = x100[1:30,]
#
#  Compute the t stat for each column
#
tv30.0.1=apply(x30,2,tst,mn=0.1)
tv100.0.1=apply(x100,2,tst,mn=0.1)
#
#  Preliminary plot with 80 bars in histogram
#
hist(tv30.0.1,breaks=80,prob=T)
curve(dt(x,df=29),-15,5,add=T,col='red')
#
#  It is obvious that the t density goes
#   too high and misses lots of the left tail
#
#  Remake plot with more bars,
#  Only plot from -15 to 5 to see central shape better
#
hist(tv30.0.1,breaks=500,prob=T,xlim=c(-15,5))
curve(dt(x,df=29,n=200),-10,5,add=T,col='red')
#
# You can see the t-distribution misses the shape badly.


#
#  Now for n=100, df =99
#
hist(tv100.0.1,breaks=80,prob=T)
curve(dt(x,df=99),-15,5,add=T,col='red')
#
#  Approximation still pretty poor
#
#  Remake plot with more bars,
#
hist(tv100.0.1,breaks=200,prob=T,xlim=c(-15,5))
curve(dt(x,df=99),-10,5,add=T,col='red')
#
# Clearly t-statistic does not have t dist.
```

```
#
#  Repeat with shape = 1
#

x100 = matrix(rgamma(1000000,shape=1),nrow=100)
x30 = x100[1:30,]
tv30.1=apply(x30,2,tst,mn=1)
tv100.1=apply(x100,2,tst,mn=1)

hist(tv30.1,breaks=80,prob=T)
curve(dt(x,df=29),-8,3,add=T,col='red')

#
# Approximation is better but not good


#
#  Now for n=100, df =99
#
hist(tv100.1,breaks=80,prob=T)
curve(dt(x,df=99),-4,4,add=T,col='red')
#
#  Approximation better but
#   density too low on left
#   and too high on right.
#
#
#  Shape = 10

x100 = matrix(rgamma(1000000,shape=10),nrow=100)
x30 = x100[1:30,]
tv30.10=apply(x30,2,tst,mn=10)
tv100.10=apply(x100,2,tst,mn=10)

hist(tv30.10,breaks=80,prob=T)
curve(dt(x,df=29),-4,4,add=T,col='red')
```

```
#
# Approximation is better but not good


#
#  Now for n=100, df =99
#
hist(tv100.10,breaks=80,prob=T)
curve(dt(x,df=99),-4,4,add=T,col='red')
#
#  By now approximation is getting much better
#

#  Shape = 100

x100 = matrix(rgamma(1000000,shape=100),nrow=100)
x30 = x100[1:30,]
tv30.100=apply(x30,2,tst,mn=100)
tv100.100=apply(x100,2,tst,mn=100)

hist(tv30.100,breaks=80,prob=T)
curve(dt(x,df=29),-4,4,add=T,col='red')

hist(tv100.100,breaks=80,prob=T)
curve(dt(x,df=99),-4,4,add=T,col='red')
#
#  t approximation quite ok even for n=30
#
```